ENFORMATIKA
Volume 11 March 2006
ISBN 1503-5313
ISBN 975-00880-1-9

Honorary Editor & Chairman
Pashayev, A. M.
National Academy of Aviation, AZ

Conference Co-Chairmen
Abiyev, R. H.
North East University, CY
Sadiqov, R. A.
National Academy of Aviation, AZ
Sherry, A. M.
Institute of Management
Technology, IN

Editor-in-Chief
Ardal, C.
National Academy of Aviation, AZ

Editorial Board
Bose, T.
Udah State University, USA
Damasevicius, R.
Kaunas University of Technology, LT
Dehmer, M.
Technische Universitat Darmstadt, DE
Frattolillo, F.
University of Sannio, IT
Hau, S.
Curtin University of Technology, AU
Kotsiantis, S.
University of Patras, GR
Mitra, A.
Indian Institute of Technology, IN
Paprzycki, M.
SWPS, PL
Seda, M.
Hml University of Technology, CZ
Sherry, A. M.
Institute of Management
Technology, IN
Tuton, M. J. C.
University of Granada, ES
Yang, L. T.
St. Francis Xavier University, CA

Published by
World Informatica Society
http://www.enformatika.org
e-mail: info@enformatika.org
Tel/Fax: +90 266 218 07 09
Barbares Mah Haber Sitesi B3-Bil 9
Canakkale 17020 Turkey

PREFACE

Dear Distinguished Delegates and Guests.

We appreciate the opportunity to welcome our distinguished delegates and guests at the twelfth International Conference on Computer Science (ICCS'06), held on March 29-31, 2006 in Vienna, Austria. This proceedings records the fully refereed papers presented at the conference. The main conference tracks are communication technology, artificial neural networks, pattern analysis, software & systems engineering.

The conference aims to bring together researchers, scientists, engineers, computer users, and students to exchange and share their experiences, new ideas, and research results about all aspects of computer science and engineering, and discuss the practical challenges encountered and the solutions adopted. The main goal of these events is to provide international scientific forums for exchange of new ideas in a number of fields that interest in-depth through discussions with their peers from around the world. The model used to form these research conferences facilitates communication among researchers in different fields of computer science and engineering. Both inward research; core areas of computer science & engineering and outward research; multi-disciplinary, inter-disciplinary, and applications will be covered during these events.

The conference has gathered submissions related to all aspects of the major conference themes and tracks. Both technical research and survey overview papers were solicited, as well as descriptions of working systems, position statements and reports on work in progress. All the submitted papers in this proceedings have been peer reviewed by at least three reviewers drawn from the scientific committee, external reviewers and editorial board depending on the subject matter of the paper. Reviewing and initial selection were undertaken electronically. After the rigorous peer-review process, the submitted papers were selected on the basis of originality, significance, and clarity for the purpose of the conference. The selected papers and additional late-breaking contributions to be presented as lectures will make an exiting technical program. The conference program is extremely rich, featuring high-impact presentations.

Besides its formal technical sessions the conference included invited plenary talks as well. The conference features a range of contributions by distinguished invited speakers drawn from academia. The invited speakers address significant recent applications of computer science methods and technologies, as well as important academic advances serving to enhance their potency and widen their applicability. The coverage of the whole domain of computer science & engineering and technologies is truly exceptional for the conference. The high quality of the program - guaranteed by the presence of an unparalleled number of internationally recognized top experts - can be assessed when reading the contents of the program.

The conference will therefore be a unique event, where attendees will be able to appreciate the latest results in their field of expertise, and to acquire additional knowledge in other fields. Attendees arriving as specialist experts will thus have the opportunity to leave the conference as enriched scientists. be they from academia or industry.

The program has been structured to favor interactions among attendees coming from many diverse horizons, scientifically, geographically, from academia and from industry. Included in this will to favor interactions are social events at prestigious sites.

Vienna possesses the right mix of qualities for favoring the success of the conference:
A wide diversity of computing and communications activities, both in industry and in research or higher education.
A strong involvement of local bodies.
Dynamic universities gathering thousands of students.
A friendly life style and a mild climate at that time of the year.
A wide variety of places of interest: natural, historical and cultural sites, not to mention high quality regional gastronomy.

We are grateful to all those who have contributed to the success of ICQS'06. We hope that all participants and other interested readers benefit scientifically from this proceedings and also find it stimulating in the process. Finally, we would like to wish you success in your technical presentations and social networking.

Enjoy during a week in Vienna!
With my warmest regards,

Ardal, C., Ed.
March 29-31, 2006
Vienna, Austria
CONTENTS

A New Hybrid RMN Image Segmentation Algorithm
Abdelouahab Moussaoui, Nabila Ferahtia, and Victor Chen
1

New Fuzzy Clustering Algorithm Applied to RMN Image Segmentation
Nabila Ferahtia, Abdelouahab Moussaoui, Khier Bennabha, and Victor Chen
9

Artificial Intelligence Techniques applied to Biomedical Patterns
Giovanni Luca Masala
14

Massive Lesions Classification using Features based on Morphological Lesion Differences
29

Robust Image Transmission Over Time-varying Channels Using Hierarchical Joint Source Channel Coding
Hatem Elmedchieh, Noureddine, Hamdi, and Ammar Bouallègue
25

Evaluation of Mixed-Mode Stress Intensity Factor by Digital Image Correlation and Intelligent Hybrid Method
E. Machado, and H. Vanada
29

A Self Configuring System for Object Recognition in Color Images
Michela Lecca
35

An Optical Flow Based Segmentation Method for Objects Extraction
C. Lodato, and S. Lopes
41

Enhancing Multi-Frame Images Using Self-Delaying Dynamic Networks
Lewis E. Hilel, Honghai Liu and David J. Brown
47

Performance Evaluation of Neural Network Prediction for Data Prefetching in Embedded Applications
Sofien Chouhou, Mohamed Chouhou, and Omar Hammami
53

Combined Sewer Overflow Forecasting with Feed-forward Back-propagation Artificial Neural Network
Ashael K. Fernando, Xiuqun Zhang, and Peter F. Kinley
58

Application of Feed Forward Neural Networks in Modeling and Control of a Fed-Batch Crystallization Process
Petra Georgieva, and Sebastião Veve de Jesus
65

An ensemble of Weighted Support Vector Machines for Ordinal Regression
Willem Waegeman, and Luc Boullart
71

Neural Network Imputation In Complex Survey Design
Sofia R. Amer
76

Solving Partially Monotone Problems with Neural Networks
Marina Velikova,Henrie Daniels, and Jd Van Gelder
82

Chateau Wines Classification based only on Aroma Information
Nicolas B. Belmont, Manuel A. Duarte-Mermeud, Victor A. Soto, Sebastián A. Salah, and Matías A. Bustos
88

Neuro-Hybrid Models for Automotive System Identification
Ventura Issanaw
94

Vision-based Network System for Industrial Applications
Taweep Sinpat, Arjun Namisonman, and Pitaya Tipsoowarn
98

Trajectory-Based Modified Policy Iteration
R. Sharma, and M. Gopal
103

2D & 3D Finite Element Method Packages of CIM Tool for Engineering PDE Problems
Chun K. Ahn, Jung Ilm Park, and Woo Hyun Kwon
109

Feasibility of the Evolutionary Algorithm using Different Behaviours of the Mutation Rate to Design Simple Digital Logic Circuits
Konstantin Musenov, Emanuele Stomeo, and Tatiana Kalmanova
114

Hybrid Markov Game Controller Design Algorithms for Nonlinear Systems
R. Sharma and M. Gopal
118

UPA-based Systems for Evolvable Hardware
Ovalle Lambe, Tatiana Kalmanova, and Emanuele Stomeo
123

A Framework for Product Development Process including HW and SW Components
Markku Dr. and Geongsok Choi
130

The Auto-Tuning PID Controller for Interacting Water Level Process
Narakan Thamkarn, Tienchat Sukrit, Arjun Namisonman, Sapan Gutpanich, and Kitti Tiraseeth
134

Measuring Process Component Design when Achieving Managerial Goals
Ekong, Upamanevar, and Twitte Nawongse
139

An Enhanced Cryptanalytic Attack on Knapsack Cipher using Genetic Algorithm
Deosam Pregg, Aditya Shastri and D.C. Agrawal
145

A Method of Authentication for Quantum Networks
Stefan Ross
149

Computer Verification in Cryptography
Markus Kaiser, and Johannes Buchmann
155
Layout Based Spam Filtering
Claudia N. Musat
101

Using Ontology Search in the Design of Class Diagram from Business Process Model
Waranart Kongsawat and Tawit Sennongs
er
105

Bandwidth Optimization through Dynamic Routing in ATM Networks: Genetic Algorithm & Tabu Search Approach
Susret Routray, A. M. Sherry, and B. R. Reddy
117

QoS Expectations in IP Networks: A Practical View
N. Arrazalaga, A. Salterian, M. Dominguez, and J. Alvaro
176

PetriNets Manipulation to Reduce Roaming Duration: Criterion to Improve Handoff Management
Hossam el-Enen Mostafa and Pavel Cicak
182

System Concept for Low Analog Complexity and High-IF Superposition Heterodyne Receivers
Marko Mirland and Hans-Joachim Jentschel
188

Packet Losses Interpretation in Mobile Internet
Hossam el-Enen Mostafa and Pavel Cicak
194

Models to Customize Web Service Discovery: Result using Static and Dynamic Parameters
Keo Leong Tan, Cheng-Sian Lee, and Han-Na Chua
198

Overload Control in a SIP Signaling Network
Masatada Ohta
205

PAPR Reduction Method for OFDM Signal by Using Dummy Sub-carriers
Pitsi Boonsawang, Arjun Namraron, Tawit Paungma and Hideo Kobayashi
211

Combinatorial Approach to Reliability Evaluation of Network with Unreliable Nodes and Unreliable Edges
Y. Shpgung
216

Downlink Scheduling and Radio Resource Allocation in Adaptive OFDMA Wireless Communication Systems for
User-Individual QoS
Liu Yuhua, Wang Chuming, Yin Changshhuai, and Yue Guangxin
221

Bandwidth Allocation in Mobile ATM Cellular Networks
Khaja Kamaluddin and Mohammed Youssoof
226

Liu Yuhua, Zhang Lihua, Yin Changshhuai, and Yue Guangxin
231

Bandwidth Allocation for A3R Service in Cellular Networks
Khaja Kamaluddin and Mohammed Youssoof
237

A Novel FFT-Based Frequency Offset Estimator for OFDM Systems
Mohd. Masoom, Mehrdad Arefianpoor, and Sayed Ali Hasham
240

Impact of Metallic Furniture on UWB Channel Statistical Characteristics by BEI
Yu-Shun Chen, Chen-Ching Chu, Chung-Hsin Huang, and Chien-Hung Chen
244

Electromagnetic Imaging of Inhomogeneous Dielectric Cylinders Buried in a Slab Medium by TE Wave Illumination
Chung-Hsin Huang, Chien-Ching Chu, and Chun Jen Lin
248

T-DOF PTD Controller Design using Characteristic Ratio Assignment Method for Quadruple Tank Process
Tanchai Saksri, Vutha Srithacaveeprajit, Arjun Namraron, Priva Kangrattana, and Thongchai Wirawatweemart
252

Reducing Cognitive Load in Learning Computer Programming
Mohammed Youssoof, Mohd. Sapiyan, and Khaja Kamaluddin
259

Decision Support System “Crop-9-DSS” for Identified Crops
Ganesan L
263

A Proposed Framework for Visualization to Teach Computer Science
Mohammed Youssoof, Mohd. Sapiyan, and Khaja Kamaluddin
266

Words Reordering based on Statistical Language Model
Theo. Athanaselis, Ntelios Bakamidis, and Ioannis Dologlou
270

Automatic Recognition of Emotionally Coloured Speech
Theo. Athanaselis, Ntelios Bakamidis, and Ioannis Dologlou
274

Data Preprocessing for Supervised Learning
S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas
277

Forecasting Fraudulent Financial Statements using Data Mining
S. Kotsiantis, E. Koumanakos, D. Tsilepsis, and V. Tsamakis
285

An Analysis of Blackouts for Electric Power Transmission Systems
Karmanitis Ioannis and Orfandis Konstantinos
289

On Problem of Parameters Identification of Dynamic Object
Kamal Akele and Cemal Aral
295

Approximation Algorithm for the Shortest Approximate Common Superstring Problem
J. S. Rebai and M. Elhamri
301

Towards Real-time Data Warehousing: Practical Approaches in Fraud Detection and Prevention
Thu Manh Nguyen
307
Chilean Wines Classification based only on Aroma Information

Nicolás H. Beltrán, Manuel A. Duarte-Mermoud, Víctor A. Soto, Sebastián A. Salah, and Matías A. Bustos

Abstract—Results of Chilean wine classification based on the information provided by an electronic nose are reported in this paper. The classification scheme consists of two parts; in the first stage, Principal Component Analysis is used as feature extraction method to reduce the dimensionality of the original information. Then, Radial Basis Functions Neural Networks is used as pattern recognition technique to perform the classification. The objective of this study is to classify different Cabernet Sauvignon, Merlot and Carmenere wine samples from different years, valleys and vineyards of Chile.

Keywords—Feature extraction techniques, Pattern recognition techniques, Principal component analysis, Radial basis functions neural networks, Wine classification.

I. INTRODUCTION

During the last decade several papers have been written concerning wine classification using information supplied by an electronic nose.

In [1] an aromatic classification of three wines of the same variety but different years (1995, 1996 and 1997) is presented. The input data for classification is obtained from an electronic nose [2] based on six sensors of conducting polymers. For classification purposes a Multilayer Perceptron (MLP) trained with the backpropagation algorithm (BP) [3] and a Time Delay Neural Networks (TDNN) trained with the Levenberg-Marquardt algorithm [3] were used.

In [4] wine classification is done using a NN with data provided by an electronic nose built by the authors using sensor commercially available. These sensors are of tin oxide and use the principle of resistance variations due to the adsorption of gas molecules on its surface.

Recently, in [5] an electronic nose based on metal oxide semiconductor thin-film sensors has been used to characterize and classify four types of Spanish red wines of the same variety of grapes. Principal component analysis (PCA) and probabilistic neural network (PNN) were used obtaining good results.

In this paper a wine classification methodology based on neural techniques is proposed, using as input data gas chromatograms of Chilean red wines of the varieties Cabernet Sauvignon, Merlot and Carmenere. The gas chromatograms are provided by an electronic nose model Fast GC Analyzer 7100 built by Electronic Sensor Technology. The sensor used by this instrument is of the type surface acoustic wave (SAW).

The first stage of the proposed methodology is concerned with the dimension reduction of the patterns preserving the original information from classification viewpoint. This is done using the feature extraction method Principal Component Analysis (PCA) [6]. Once the dimension of the input data has been reduced the information pass to a classification stage where a technique based on radial basis functions neural networks (RBFNN) [6, 7] is used. The general scheme of the proposed methodology is shown in Fig. 1.

In Section II a brief description of the feature extraction and the classification techniques used in this study are presented. Section III is devoted to explain the methodology used in this work. The results obtained together with a discussion are presented in Section IV. Finally some conclusions are drawn in Section V.

Fig. 1 Block diagram of the methodology proposed in the paper for wine classification

II. BRIEF DESCRIPTION OF THE FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES

In this section we present a brief description of the feature extraction and classification (pattern recognition) techniques used in this study. The description of each technique is made only for completeness and for a more detailed description the reader is referred to the cited references.

A. Principal Components Analysis (PCA)

In the Principal Components Analysis (PCA) method the main idea is to transform the original feature space into one in which the data is not correlated. This new space is obtained by projecting the original data onto a set of orthogonal axis in which the variance of the input data is maximized. This technique can be summarized in the following theorem of the principal component analysis [6].
**Fundamental Theorem for PCA**

Given a set of variables \( x_i \in \mathbb{R}^n \) for \( i=1,2,...,n \) with a non singular covariance matrix \( \Sigma \in \mathbb{R}^{n \times n} \), it is always possible to define a subset of non correlated variables \( y_i \in \mathbb{R}^p \) for \( i=1,2,...,n \) by means of a linear transformation \( W \), corresponding to a rigid rotation, whose columns are eigen-vectors of \( \Sigma \). The covariance matrix of the new set of variables \( \Sigma \in \mathbb{R}^{p \times p} \) is diagonal and contains the eigen-values \( \lambda_i \) for \( i=1,2,...,n \) associated to the eigen-vectors which are columns of \( W \).

Form the previous Theorem the eigen-values \( \lambda_i \) can be seen as the variance of the patterns in the transformed space, which are related to the range of the patterns of each axis of this space. On the other hand, the eigen-vectors \( \phi_i \) associated to the eigen-values \( \lambda_i \) determine the direction of maximum variance. Thus selecting a subset of eigen-vectors a rotation will be performed that align the transformed axes with the direction of the maximum variance of data. The dimension reduction will be determined by the size of the set of eigen-vectors chosen.

In Section III E a detailed explanation is presented, as to how the PCA methodology is used in this particular case to reduce the dimension of the input data to the classifier.

**B. Radial-Basis Functions Neural Networks (RBFNN)**

When Pattern recognition techniques based on neural networks have shown a great behavior for a wide range of applications [6, 8] and they are very attractive since a minimum knowledge on the patterns is required. The radial-basis functions neural networks (RBFNN) constitute the main alternative to the multi-layer perceptron (MLP) for data interpolation and pattern classification problems. They are characterized since instead of using a linear activation function they use functions with symmetry around a center \( c \) in the \( n \)-dimensional space of the input patterns.

Each neuron in a RBFNN corresponds to a region in the \( n \) dimensional input space with center \( c \). The activation level of a neuron in a RBFNN to an input \( x \) is a function of the Euclidean distance between \( x \) and the center \( c \) of the neuron. The output of the neuron in a RBFNN is given by the general equation

\[
y_i(x) = \sum_{j=1}^{m} w_{ij} \phi_j(x)
\]

where \( \phi_j(x) \) are the RBF and \( w_{ij} \) are the weights in the output layer (See Fig. 3). The basis functions \( \phi_j(x) \) can be interpreted as the a posteriori probabilities \( p(j|x) \) indicating the presence of certain characteristics in the input space. Similarly, the weights \( w_{ij} \) can be interpreted as the a posteriori probabilities \( p(C_i / j) \) of the members of a class given certain input characteristics. That is the reason why it is natural to apply RBFNN to pattern classification problems [8]. The basic unit of a RBFNN is shown in Fig. 2.

![Fig. 2 Mathematical model of a neuron in a RBFNN](image)

In this study basis functions \( \phi(x) \) for one neuron is of Gaussian type

\[
\phi(x) = \exp \left( -\frac{(x - c)^2}{2\sigma^2} \right)
\]

where \( c \) determines the center and parameter \( \sigma \) determine the size of the receptive field. \( \sigma \) is also known as spread and defines the selectivity of the neuron as

\[
selectivity = \frac{1}{\sigma}
\]

A small \( \sigma \) implies a high selectivity whereas a large value of \( \sigma \) makes the neuron less selective.

A RBFN is composed of two layers one containing the RBF and a linear output layer summing the multiplication of the outputs of the RBF with the vector of weights, as seen in Fig. 3.

![Fig. 3 Two layer architecture of a RBFNN](image)

Then it is necessary to define the spread \( \sigma \) and the centers \( c \), of the neurons forming the receptive fields of the network. The usual way is to set the center \( c \) at each one of the training patterns of the problem. Thus, if we have \( p \) training patterns the network has \( p \) neurons centered at each pattern. This strategy guarantee zero error in the training set and the freedom to choose \( \sigma \) that generates a controlled spatial overlapping to guarantee a good generalization. Depending on
the computational implementation utilized. \( \sigma \) can be equal for all neurons or have different values for each unit.

The next step is to choose the weight vector \( w \in \mathbb{R}^m \). To this extent the RBFNN is evaluated at the \( \rho \) training patterns

\[
\phi_{\mu} = \phi \| x_i - x_j \| \quad \forall i, j = 1 \cdots m
\]

(4)

where \( \| \| \) corresponds to the Euclidean norm between two vectors.

We define matrix \( \Phi \) composed by all the \( \phi_{\mu} \) as the interpolation Matrix of the problem [9], from which the weights can be obtained through the relationship

\[
\Phi w = T
\]

(5)

where \( w \in \mathbb{R}^m \) is the weight vector and \( T \in \mathbb{R}^m \) is the objective vector (target) containing the desired outputs.

Then if \( \Phi \) is nonsingular the weights are obtained as

\[
w = \Phi^{-1} T
\]

(6)

The Michelli's Theorem [9] guarantee that if all vectors \( x_i \) used to compute \( \Phi \) are all different, then \( \Phi \) will be nonsingular.

III. EXPERIMENTAL SETUP

A. Electronic Nose

If the electronic nose used in the study is the model Fast GC Analyzer 7100 built by Electronic Sensor Technology [13] with surface acoustic wave (SAW) sensor. The most important operating parameters for the electronic nose are:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor</td>
<td>60°C</td>
<td></td>
</tr>
<tr>
<td>Column</td>
<td>40°C</td>
<td></td>
</tr>
<tr>
<td>Valve</td>
<td>140°C</td>
<td></td>
</tr>
<tr>
<td>Inlet</td>
<td>175°C</td>
<td></td>
</tr>
<tr>
<td>Trap</td>
<td>300°C</td>
<td></td>
</tr>
<tr>
<td>Ramp</td>
<td>10°C/s</td>
<td></td>
</tr>
<tr>
<td>Acquisition time</td>
<td>20s</td>
<td></td>
</tr>
<tr>
<td>Sampling rate</td>
<td>0.01s</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4 Electronic nose model Fast GC Analyzer 7100 from Electronic Sensor Technology

B. Database

The database used in the study is formed by 100 commercial samples of Chilean wines of the type Cabernet Sauvignon, Merlot and Carmenère. These wines belong to the vintages 1997-2003 and come from different valleys of the central part of Chile. The distribution of the samples is shown in Table II.

<table>
<thead>
<tr>
<th>Class</th>
<th>Type</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cabernet Sauvignon</td>
<td>36</td>
<td>36%</td>
</tr>
<tr>
<td>2</td>
<td>Merlot</td>
<td>44</td>
<td>44%</td>
</tr>
<tr>
<td>3</td>
<td>Carmenère</td>
<td>20</td>
<td>20%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>100%</td>
</tr>
</tbody>
</table>

The information from each sample was obtained setting the parameters shown in Table I for the instrument. 10 measurements were done for each of the 100 sample obtaining in total 1000 profiles (chromatograms).

C. Data Pre-Processing

The chromatograms obtained from the electronic nose are curves of 12 sec. of duration with a sampling period of 0.02 sec., containing 600 points in total. Fig. 5 shows a typical profile corresponding to a Cabernet Sauvignon sample properly normalized.

The normalization process was done to normalize the
amplitude in the interval [-1,1]. To this extent the maximum amplitude was used to normalize the signal according to the following relationship:

\[ x'_i = \frac{x_i}{x_{\text{max}}} \]  \hspace{1cm} (7)

where \( x_{\text{max}} \) is the maximum amplitude of all profiles.

Fig. 5 Typical normalized chromatogram for a Cabernet Sauvignon (600 points and 12 s of duration)

**D. Methodology**

In order to classify the profiles described in Section III B, the classification technique Radial Basis Functions Neural Networks (RBFNN) described in Section II B will be used. Because of the high data dimensionality, previous to the classification it was necessary to perform a feature extraction procedure of the original data using the Principal Component Analysis (PCA) technique described in Section II A.

The total database of 1000 profiles (360 Cabernet Sauvignon or Class 1, 440 Merlot or Class 2 and 200 Carménère or Class 3) was divided in two sets; one for training-validation (containing the 90% of the samples) and other for Test (containing the 10% of the samples). The sample distribution is the following:

**Training-Validation Set:** 900 profiles corresponding to 90 wine samples, 330 profiles Cabernet Sauvignon (33 samples), 390 profiles Merlot (39 samples), and 180 profiles Carménère (18 samples).

**Test Set:** 100 profiles corresponding to 10 wine samples, 30 profiles Cabernet Sauvignon (3 samples), 50 profiles Merlot (5 samples), and 20 profiles Carménère (2 samples).

The samples for each set were selected randomly and based on the proportion of the samples of different kind contained in the original database.

As a measure of the behavior and to obtain the optimal values of the parameters for each method, Cross-Validation was used \([8,10,11]\). The database was divided in \( n \) sets, using \( n-1 \) for training and the reminder for validation. The process is repeated \( n \) times so that all \( n \) sets are used once for validation.

In the training-validation it will be used cross-validation with the aim to measure the behavior and to tune the optimal parameters for each classification feature extraction method. Then each classifier will be evaluated with the Test set, using the whole training-validation set and the optimal parameters determined by cross validation. It is important to notice that the Test Set is never used in the training stage and therefore will be completely unknown go the classifier and is a good performance measure of each method.

Since there is 10 profiles for each wine sample, the size of cross validation sets will be 10 and then the training-validation base will be divided in 90 sub-sets of 10 elements, each one representing one wine sample. Thus, for each method the training is done using 890 profiles and one simulation for validation having 10 elements. The process is repeated 90 times so that each subset of 10 elements is used once to validate the method. The measure of the behavior will be the average and the standard deviation of the percentage of correct classification in validation.

Finally, once cross validation is done and the optimal parameters are found for each method, one simulation is done with the Test Set to evaluate the performance of each method when unknown samples are presented. The behavior will be measured again in terms of the average and the standard deviation of the percentage of correct classification in the test set.

**E. Feature Extraction using Principal Components Analysis (PCA)**

The central idea of the principal components analysis (PCA) is to transform the input space of the variables \( P \) onto a space \( P' \) where the data is not correlated i.e. the variance of the data is maximum. This is achieved by computing the eigen-values and the eigen-vectors of the of covariance matrix of the initial data and selecting those eigenvectors that have the largest eigen-values. These components represent the axes of the new transformed space. By projecting the initial data onto these axes the largest data variance is obtained.

The profiles can be seen as characteristic vectors belonging to \( R^{100} \) and the database as a matrix of \( 600 \times 1000 \), where the 1000 columns correspond to each column and the 600 rows to the points (that are going to be reduced).

Considering the training-validation set we have a matrix of \( 600 \times 900 \) (900 profiles (columns) of 600 points (rows)), then the covariance matrix of the training-validation set is

\[ \Sigma_X = X'X^T \]  \hspace{1cm} (8)

with \( X' \) the training-validation matrix and \( \Sigma_X \) the covariance matrix of \( X' \) of \( 600 \times 600 \). Then computing the eigen-values and the eigenvectors of \( X' \) and selecting the eigen-vectors with the largest eigen-values the principal components transformation matrix will be determined. One way of choosing the eigen-values (and the eigenvectors associated) is considering the contribution to the global variance \([8,12]\) of...
each eigen-value, \( \gamma_i \), as:

\[
\gamma_i = \lambda_i / \sum \lambda_i
\]

(9)

being \( N = 600 \) the total number of eigen-values of the covariance matrix \( X \).

\( \gamma_i \) associates to each eigen-value (and each eigen-vector or principal component) a factor of relative importance considering its contribution to the total variance.

When computing the eigen-values of the covariance matrix \( X \), these are ordered in ascending order [11,19], thus the last components are those contributing the most to the information (in terms of the covariance) whereas the first can be considered as noise and therefore disregarded. In Fig. 6 are plotted the last 25 eigen-values (of a total of 600) of the training-validation covariance matrix of 600 x 600.

![Graph of Eigen-values of the Covariance Matrix (Training-Validation)](image)

Fig. 6 Detail of the last 25 eigen-values of the training-validation covariance matrix

Fig. 6 shows only the last 25 eigen-values of the values training-validation covariance matrix and it is observed that these retain practically all the information in terms of the covariance. When computing the contribution of the last 20 eigen-values to the global covariance using (9), the contribution to the total information is 99.87% and the last 10 eigen-values contribute 99.46%. Therefore we will choose the matrix transformation composed by the 20 eigen-vectors associated to the last 20 eigen-values, generating a 600x20 matrix (the 600 rows represent the initial characteristics or points and the 20 columns the eigen-vectors or new characteristics). Multiplying each original profile by the transformation matrix a low dimension profile is obtained (dimension 20) which will be used in the classification procedure.

IV. RESULTS

As explained in Section II B, in a RBFNN is necessary to define the centers of the neurons and the parameters \( \sigma \) known as spread, which define the selectivity of the neuron. For all simulations the neurons were located at each training pattern [28], thus when cross validation is carried out the networks has 890 neurons corresponding to each profile. Recall that the NN has two layers; the first has radial basis activation functions and the second linear activation functions.

Simulations were carried out making cross validation with the training-validation set for different values of the selectivity \( \sigma \), and computing the performance. The same was done for the test set.

For this method 20 principal components containing the 99.86% of the total information of the training-validation data were considered. That is to say the data dimension is reduced from 600 to only 20 pints.

Different values of the selectivity were considered in the interval [2-9, 10]. For higher values of the selectivity the results were poor in both validation and test sets. The results obtained are presented in Table III.

<table>
<thead>
<tr>
<th>Selectivity</th>
<th>% Correct Classification in validation</th>
<th>Standard Deviation</th>
<th>% Correct Classification in test</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>39.8</td>
<td>0.4913</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>36.6</td>
<td>0.4845</td>
<td>50</td>
</tr>
<tr>
<td>0.1</td>
<td>60.8</td>
<td>0.4402</td>
<td>52</td>
</tr>
<tr>
<td>0.02</td>
<td>35.2</td>
<td>0.3557</td>
<td>63</td>
</tr>
<tr>
<td>0.01</td>
<td>53.5</td>
<td>0.3846</td>
<td>67</td>
</tr>
<tr>
<td>0.0078125</td>
<td>61.3</td>
<td>0.3641</td>
<td>65</td>
</tr>
<tr>
<td>0.00390625</td>
<td>66.1</td>
<td>0.3853</td>
<td>76</td>
</tr>
<tr>
<td>0.00195313</td>
<td>71.4</td>
<td>0.3776</td>
<td>60</td>
</tr>
</tbody>
</table>

Observing Table III it is appreciated that classification results do not reach the 80% of correct classification in the best case.

The simulations were carried out using the same computer already mentioned in the previous section and using the software Matlab 6.0, the Neural Network Toolbox and the Signal Processing Toolbox.

The average processing time for each simulation are shown in Table IV as a function of the number of principal components chosen. The average is computed over three runs performed for each simulation.
The results exposed above indicate the necessity of having a larger database with uniform distribution of the classes. Finally, it is important to point out that the results gotten in this study are auspicious and corroborate that in spite of the reduced dimension of the database it is indeed possible to classify Chilean wines according to the cepas using chromatograms coming from an electronic nose.

V. CONCLUSIONS

The classification of red Chilean wines of the type Cabernet Sauvignon, Merlot and Carménère, from different vintages and different valleys, was successfully performed based on the gas chromatograms supplied by an electronic nose.

Principal Component Analysis was used as feature extraction and Radial Basis Function Neural Networks as classification technique. The best parameters for each method were obtained from the cross validation process with the training-validation set.

The RBFNN showed a discrete performance in the training-validation set with classification rates about 71% (for $\sigma=0.00195$) and a 76% in the test set (for $\sigma=0.00391$).

For future studies it is suggested to analyze other feature extraction techniques like wavelets and other classification techniques as Support Vector Machines (SVM) currently we are working in incorporating more wine samples to the original database to complete soon 200 samples. The idea is to repeat this study with this new database these new techniques.

The results obtained in this study are promising and the first on Chilean wines using gas chromatograms supplied by an electronic nose. They provide the basis for future work on classification of Chilean wines. Other work developed by the authors on this subject can be found in [14, 15, 16].

REFERENCES